



may 28th,
CIRM

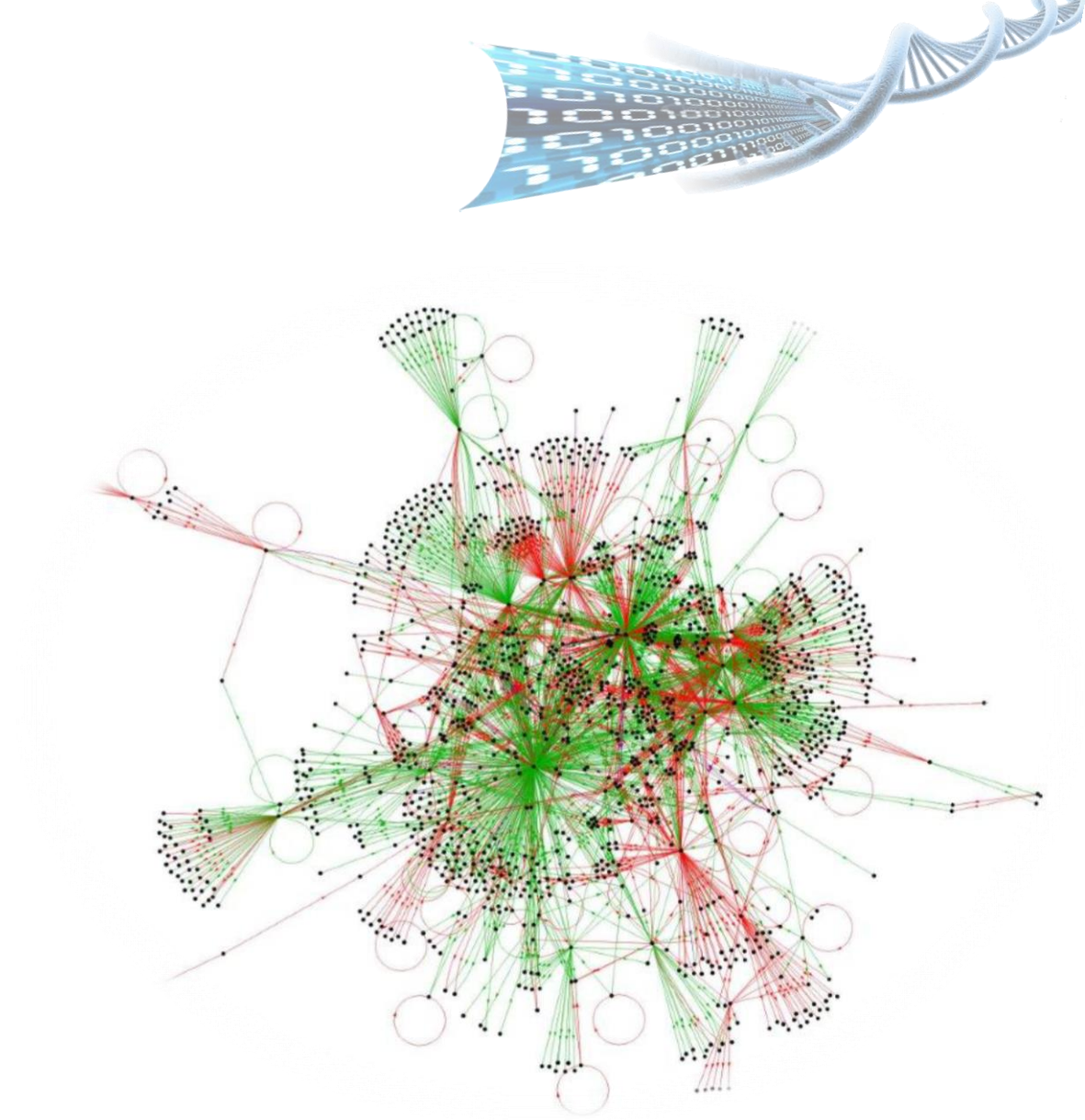
Kami :
**An annotated Knowledge representation model
for biology**

**Adrien Basso-Blandin,
Russell Harmer,
Walter Fontana**

Big mechanisms

Problematic in biology :

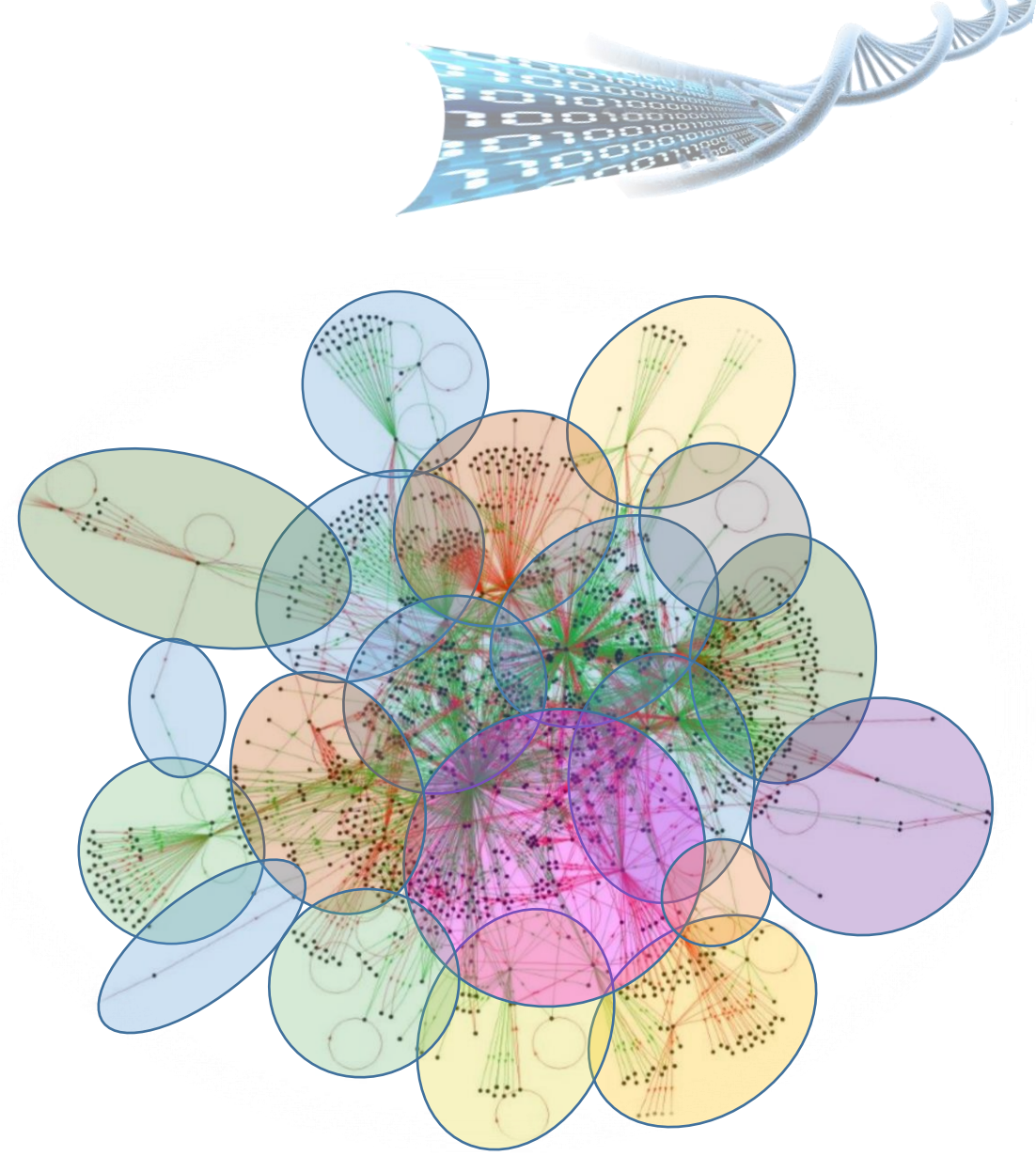
- Large and complex systems



Big mechanisms

Problematic in biology :

- Large and complex systems
- Studied piecewise
- Literatures and data fragmented



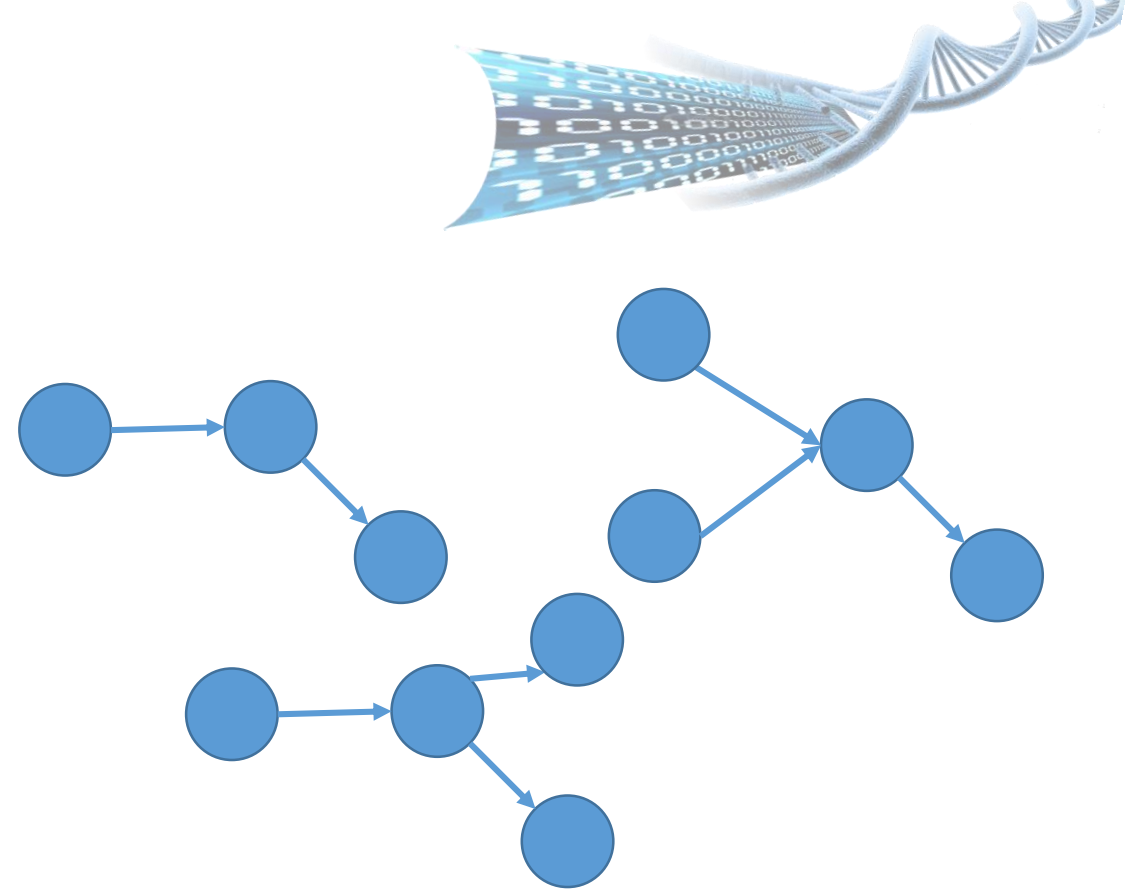
Big mechanisms

Problematic in biology :

- Large and complex systems
- Studied piecewise
- Literatures and data fragmented

Objectives :

- Aggregate knowledge into big mechanistic models



Big mechanisms

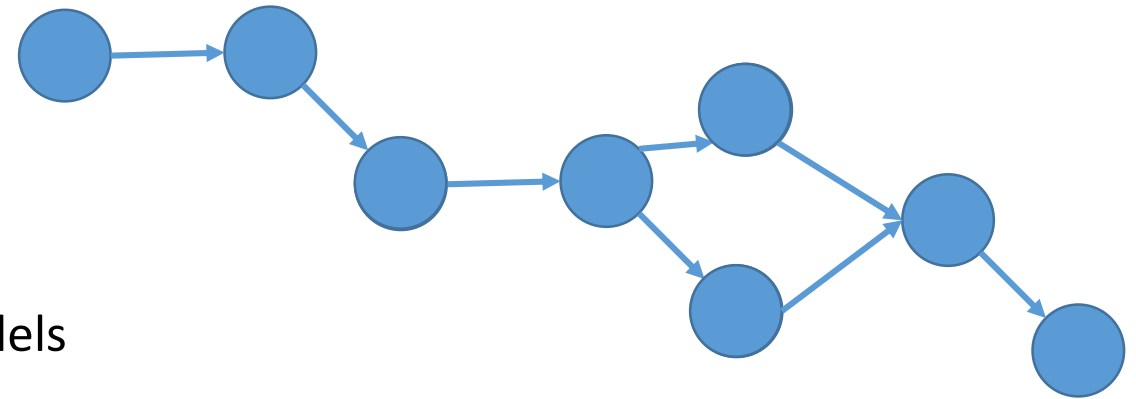


Problematic in biology :

- Large and complex systems
- Studied piecewise
- Literatures and data fragmented

Objectives :

- Aggregate knowledge into big mechanistic models



Big mechanisms

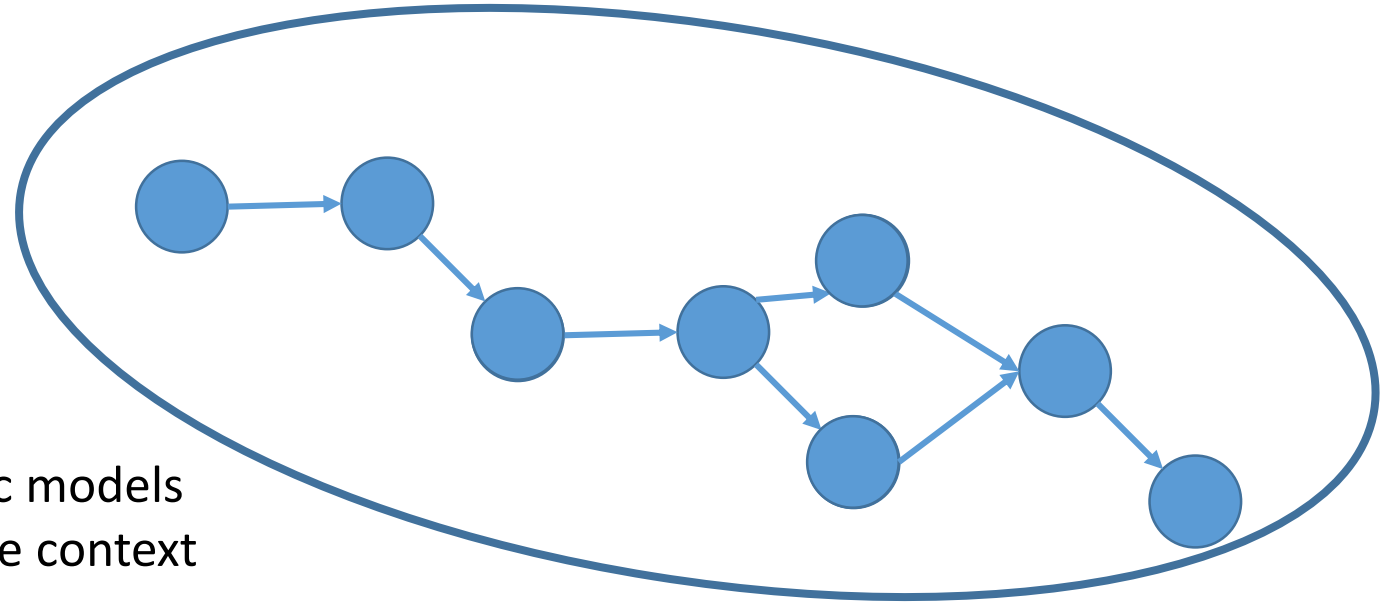


Problematic in biology :

- Large and complex systems
- Studied piecewise
- Literatures and data fragmented

Objectives :

- Aggregate knowledge into big mechanistic models
- Understand biological processes into there context



Big mechanisms

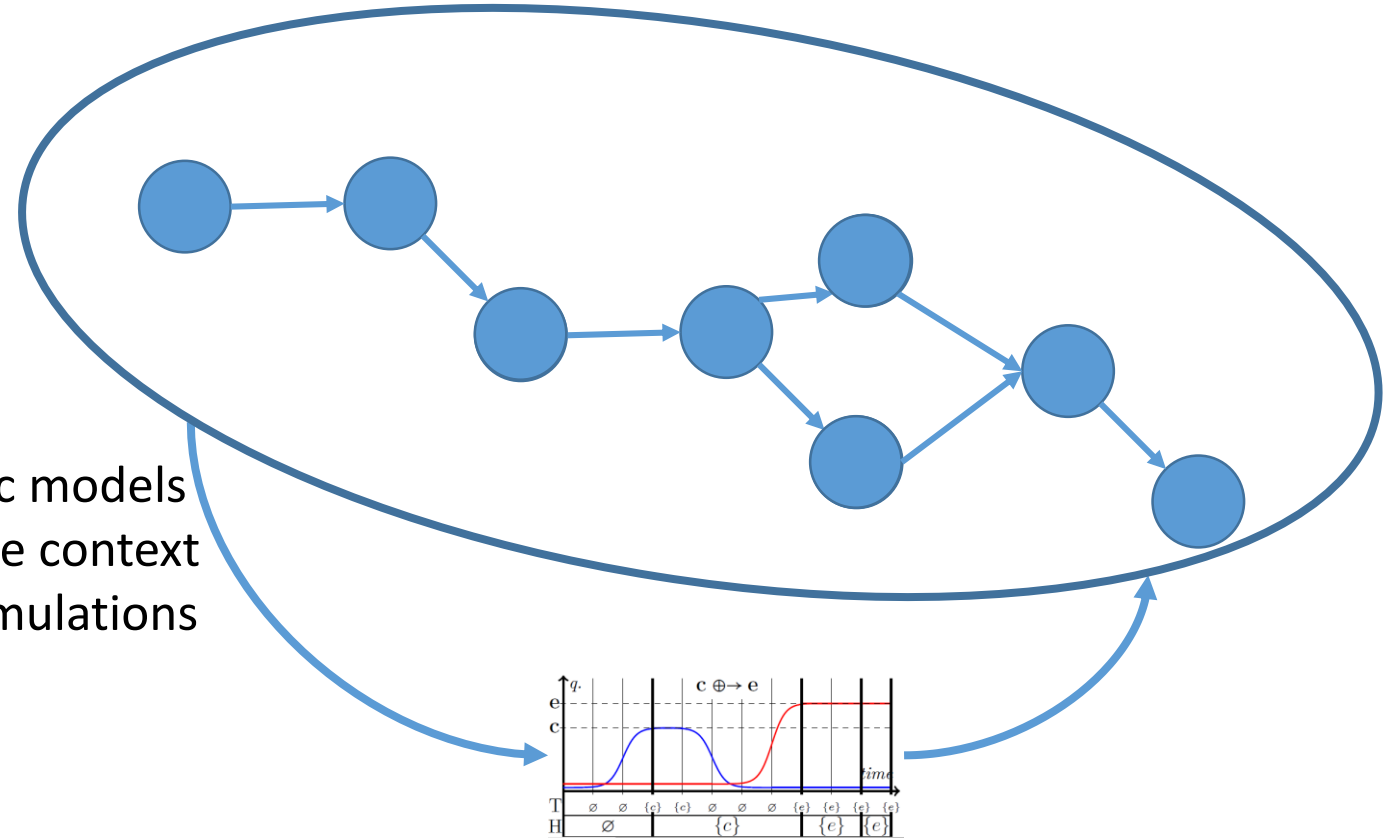


Problematic in biology :

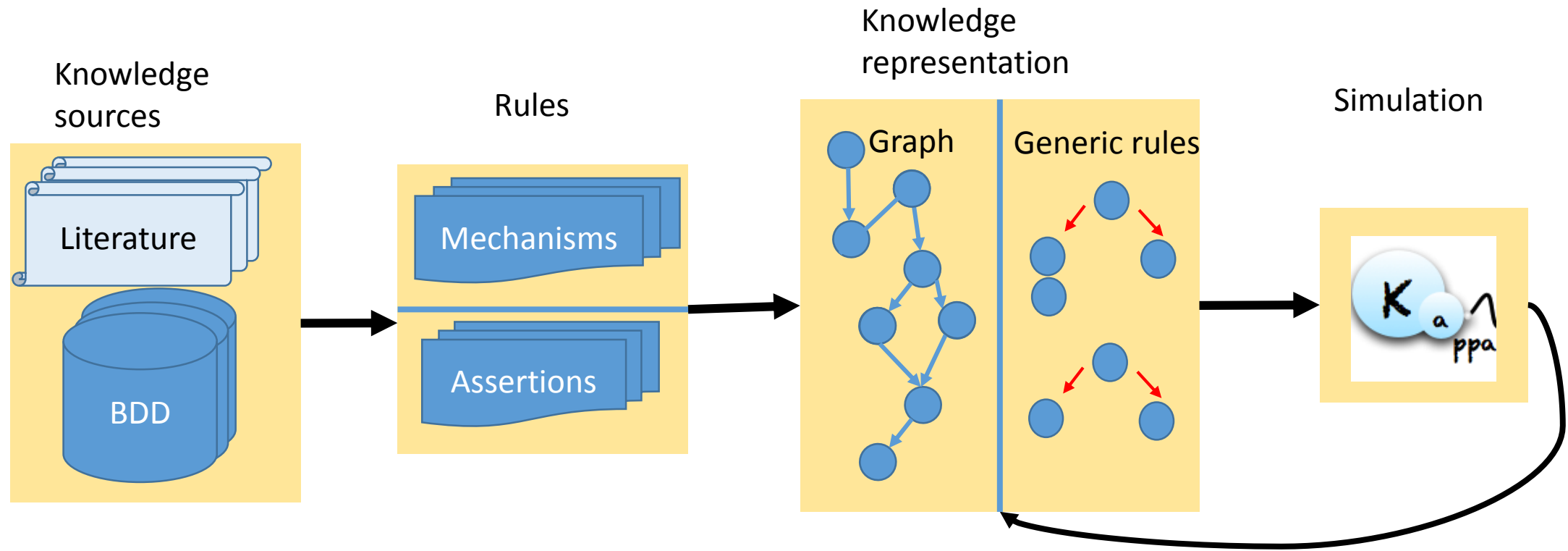
- Large and complex systems
- Studied piecewise
- Literatures and data fragmented

Objectives :

- Aggregate knowledge into big mechanistic models
- Understand biological processes into there context
- Infer new knowledge from models and simulations



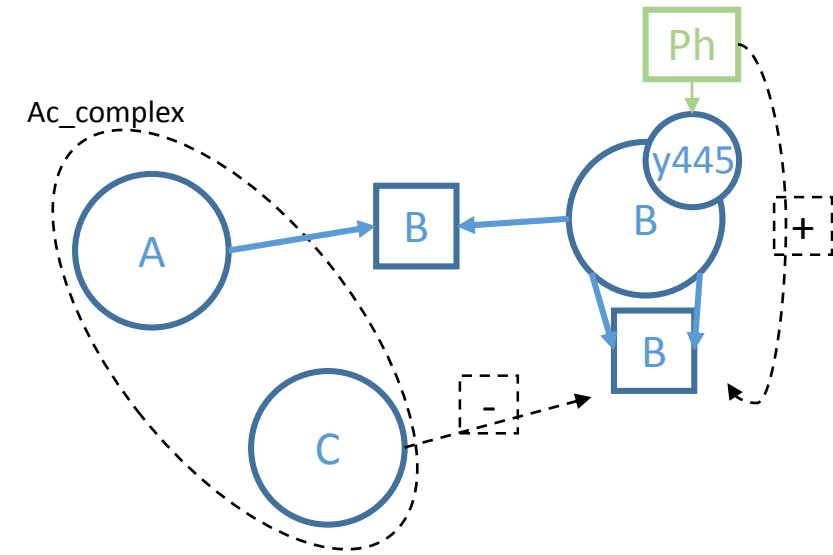
Big Mechanism



Biological tool-example



1. Protein A is an enzyme that catalyzes the phosphorylation of protein B on key residue Y445
2. Phosphorylation of B has a positive influence on its homodimerization
3. C is a known inhibitor of B homodimerization
4. C is often found in complex with A



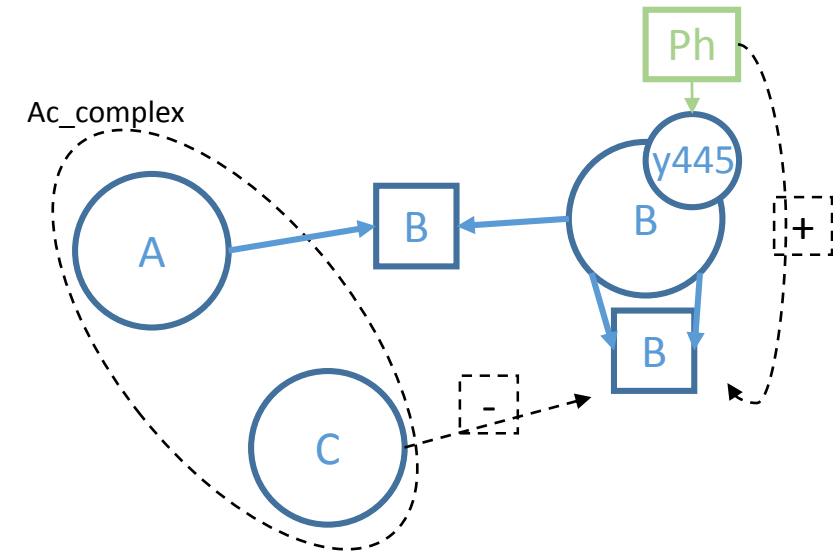
Ph : Bind(A,B)



Knowledge representation

Problematics

1. Size of the data structures
2. Nuggets aggregation
3. Knowledge are 1 level up of KR (Assertions)
4. Simulation are 1 level down of KR (Kappa)
5. User friendly GUI (for biologists)



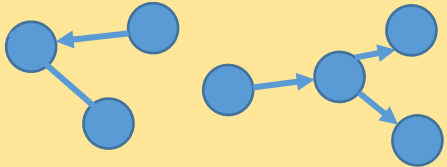
Ph : Bind(A,B)

Knowledge representation

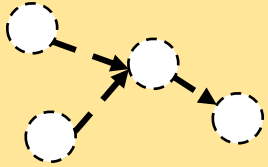


Nuggets

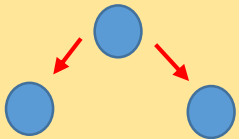
Mechanistic rules



assertion rules

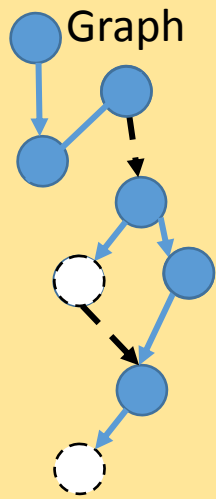


Graph rewriting rules

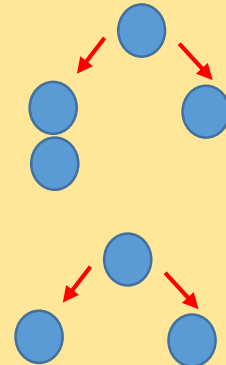


Nugget Graph

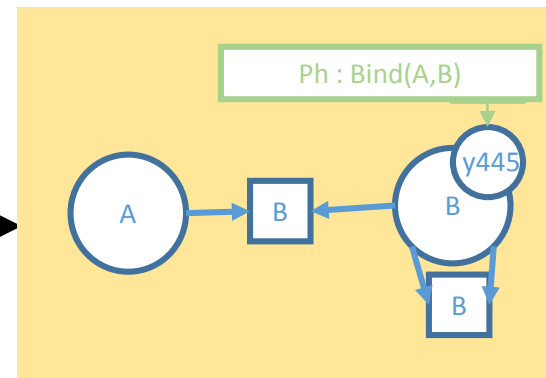
Graph



Generic rules



Logical contact graph

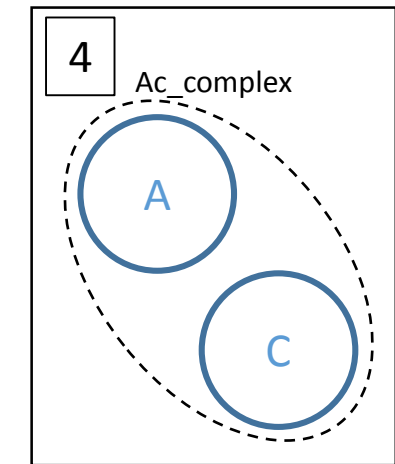
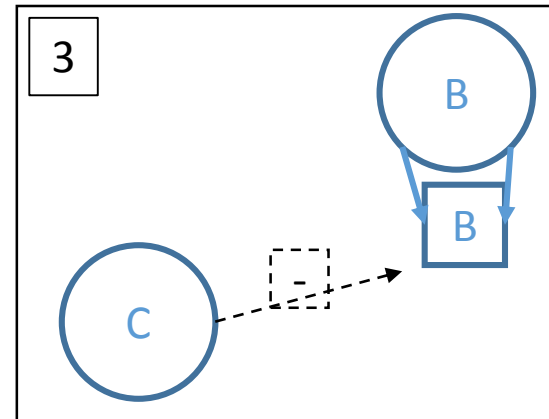
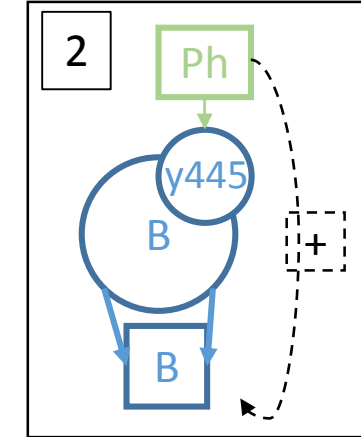
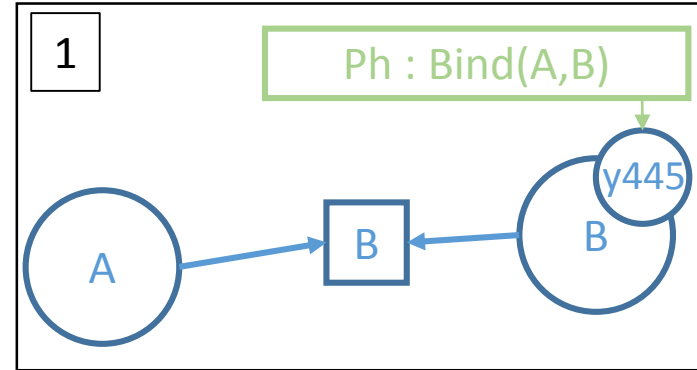


Simulation





1. Protein A is an enzyme that catalyzes the phosphorylation of protein B on key residue Y445
2. Phosphorylation of B has a positive influence on its homodimerization
3. C is a known inhibitor of B homodimerization
4. C is often found in complex with A

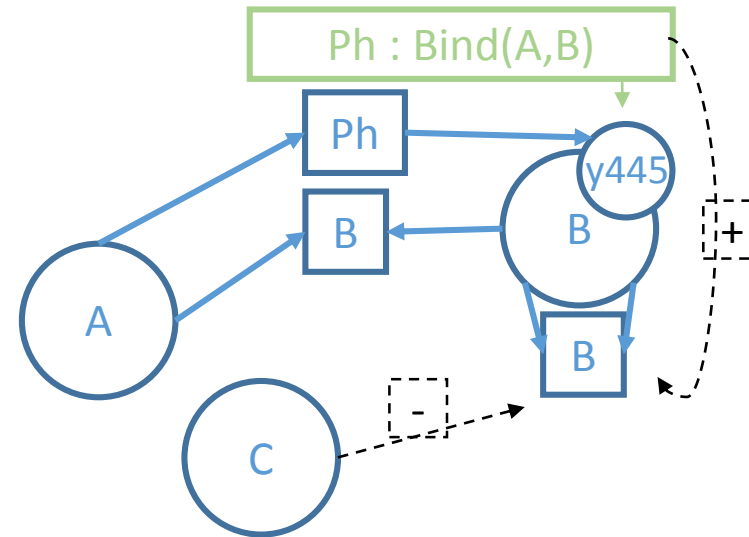


Nugget Graph



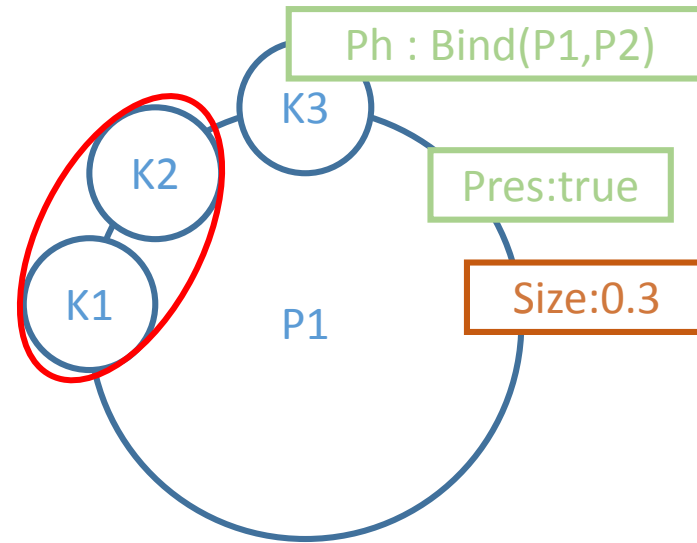
A bi(o)graph :

- Actions
- Agents
- Mechanistic edges
- Assertion edges





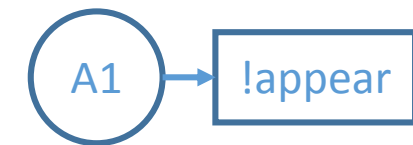
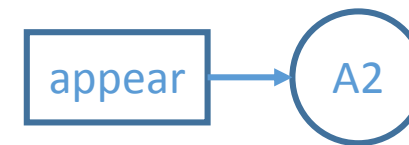
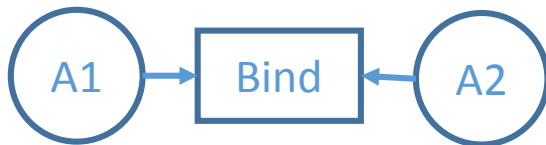
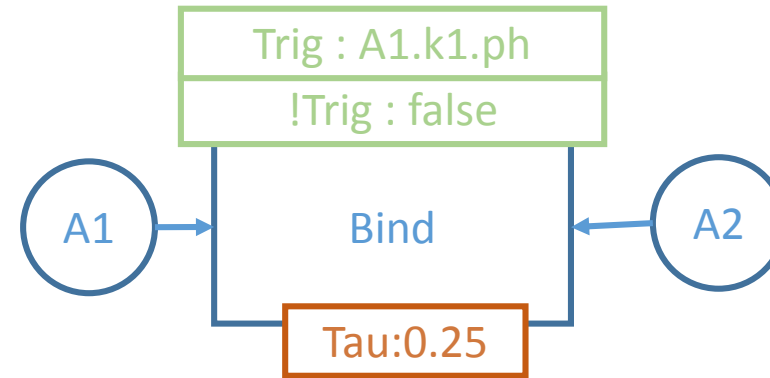
- Protein
- Key residues
- Sites
- Flags
- Attributes



Actions



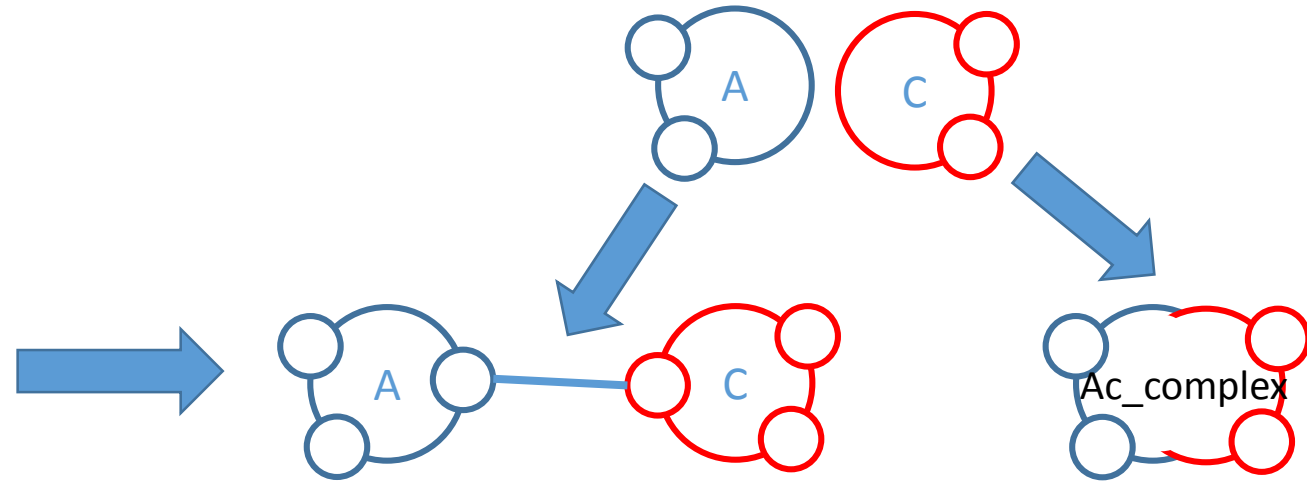
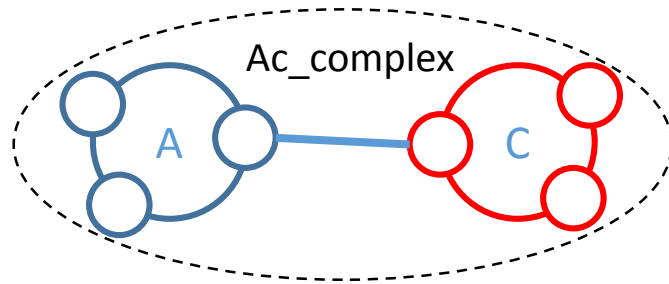
- Triggers
- Attributes
- Arity



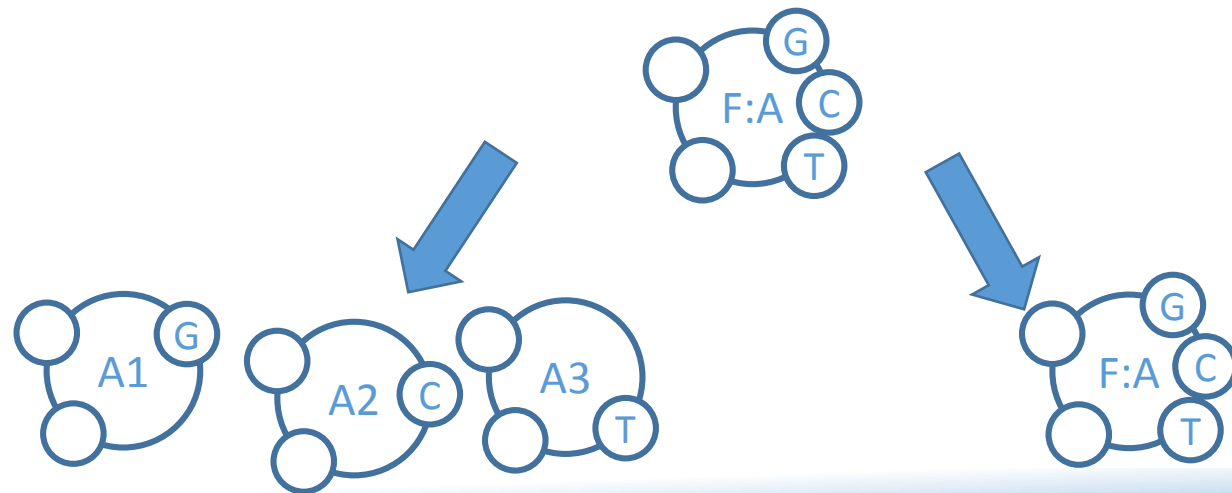
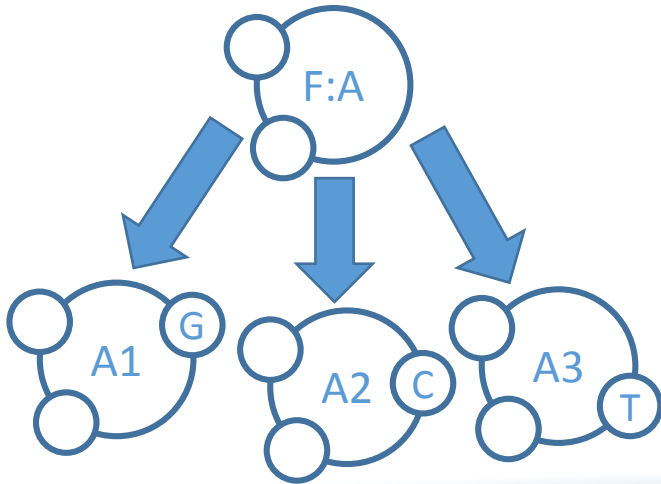
Generic rules

Example :

- Rule for complexes



- Rule for families

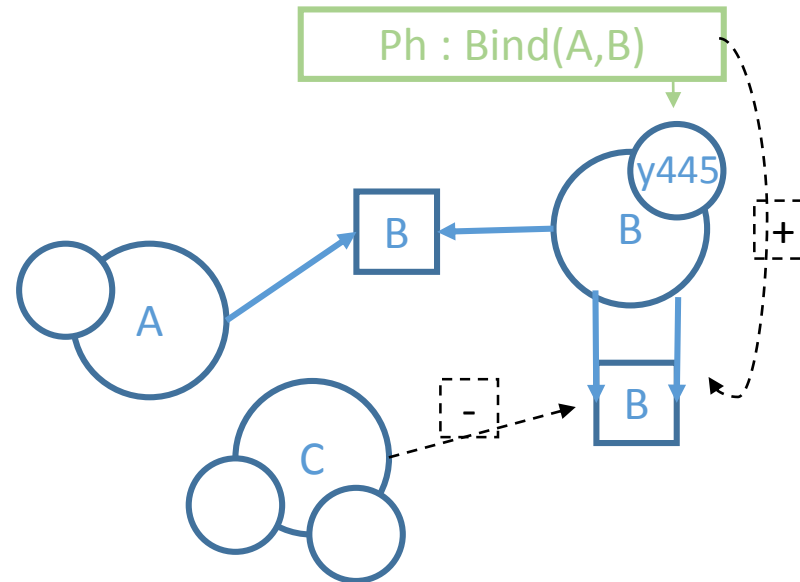


Logical contact graph



A logical graph :

- Actions
- Agents
- Mechanistic edges

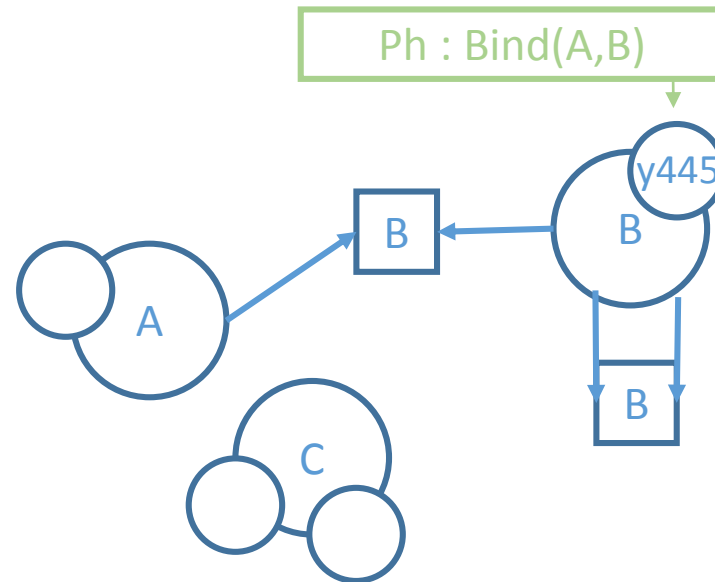


Logical contact graph



A logical graph :

- Actions
- Agents
- Mechanistic edges
- ~~Assertion edges~~

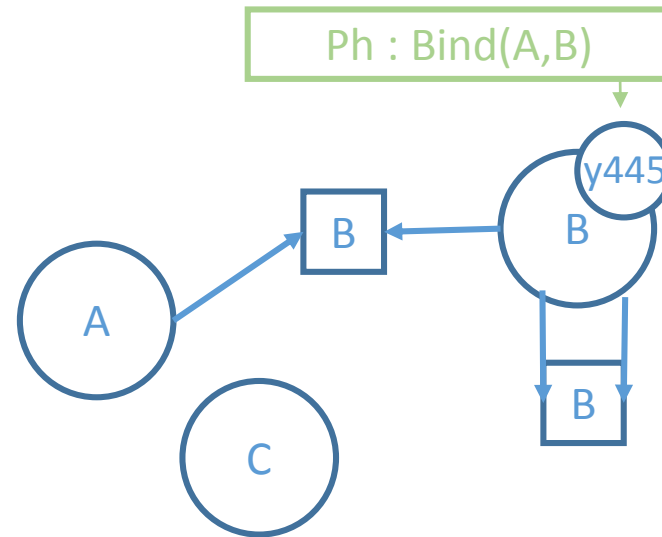


Logical contact graph



A logical graph :

- Actions
- Agents
- Mechanistic edges
- ~~Assertion edges~~
- No more useless key residues

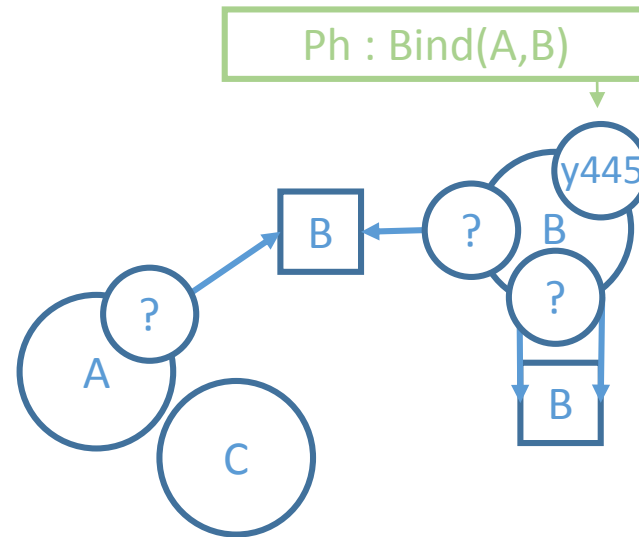


Logical contact graph



A logical graph :

- Actions
- Agents
- Mechanistic edges
- ~~Assertion edges~~
- No more useless key residues
- Virtual key residues added

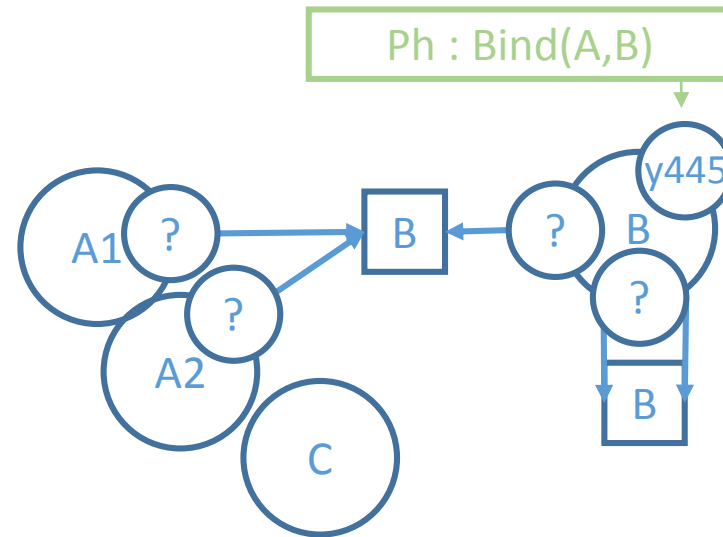


Logical contact graph



A logical graph :

- Actions
- Agents
- Mechanistic edges
- ~~Assertion edges~~
- No more useless key residues
- Virtual key residues added
- Applied generic rules

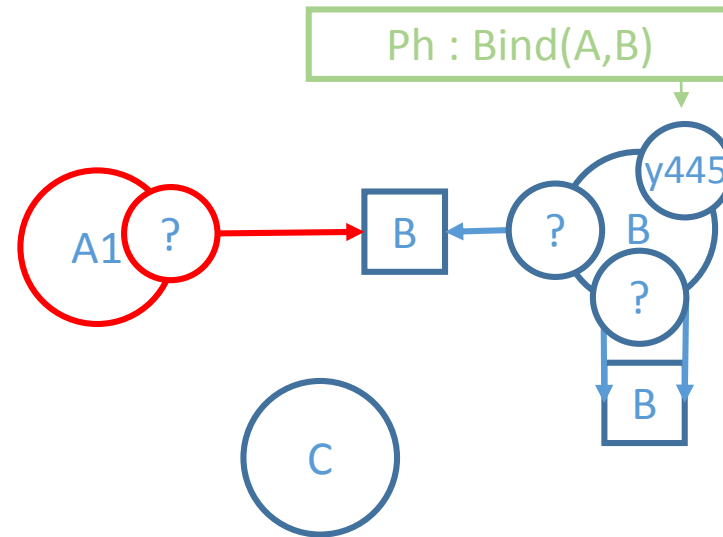


Logical contact graph



A logical graph :

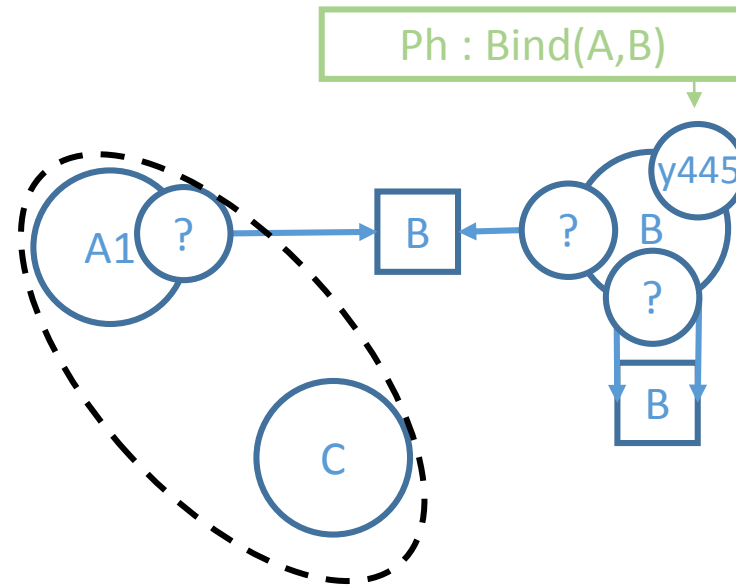
- Actions
- Agents
- Mechanistic edges
- ~~Assertion edges~~
- No more useless key residues
- Virtual key residues added
- Applied generic rules
- Instanced nugget graph



Logical contact graph



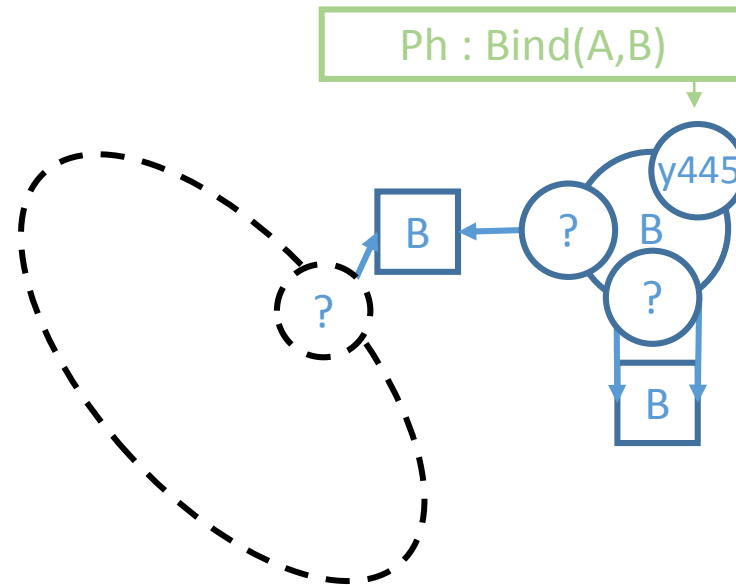
Set abstraction :



Logical contact graph



Set abstraction :

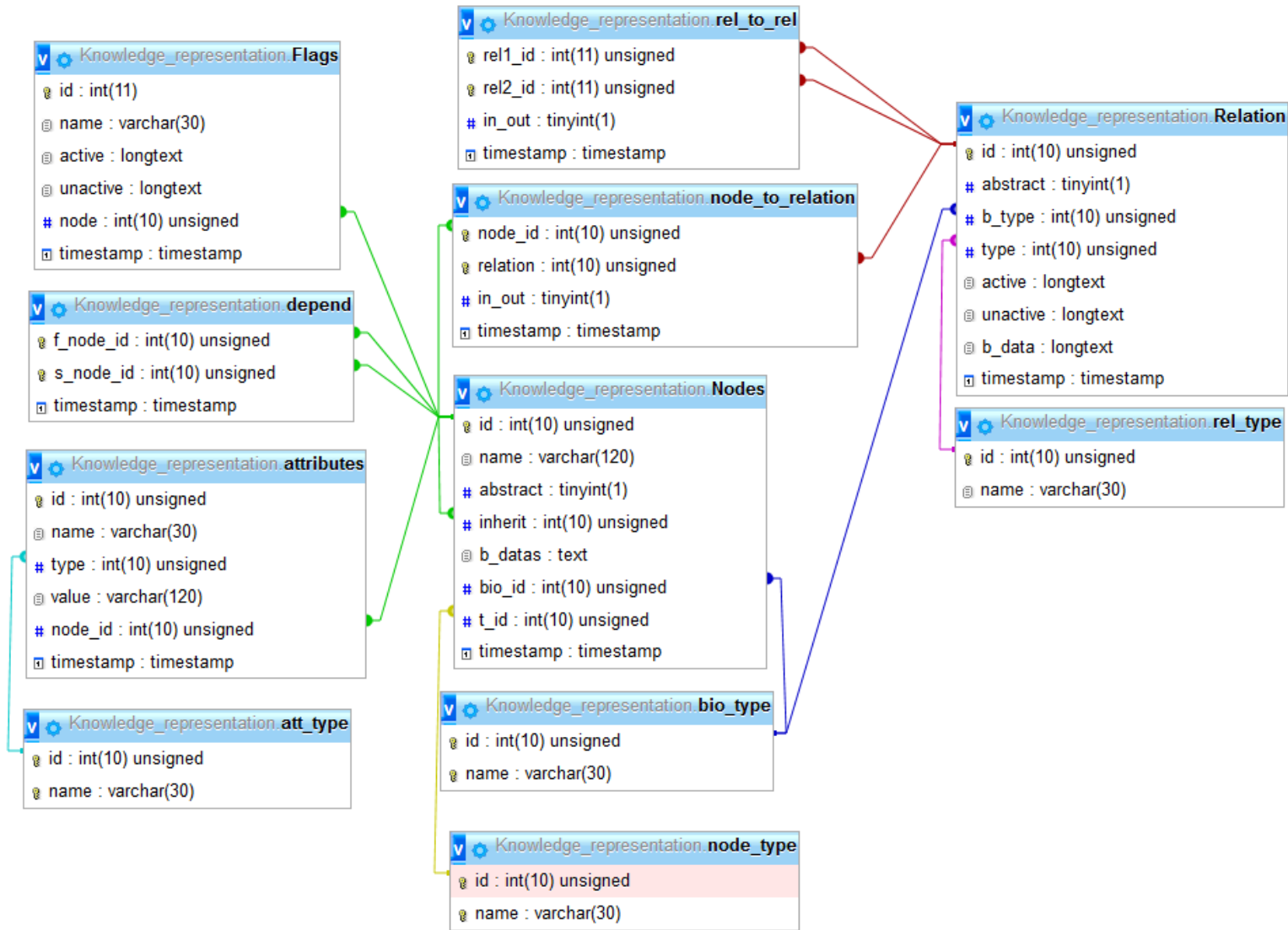




graph	::=	$(\mathcal{N}, \mathcal{R}, \mathcal{E})$ where $\mathcal{N}, \mathcal{R}, \mathcal{E}$ are respectively sets of node, actions and edges
node	::=	$(\text{string}, \text{bool}, \mathcal{T}, \mathcal{F}, S \subset \mathcal{S}, K \subset \mathcal{K}, \text{bio_t}, \text{bio_d}) \in \mathcal{A} \subset \mathcal{N}$ $ (\text{string}, \text{bool}, \mathcal{T}, \mathcal{F}, \emptyset, \emptyset, \text{bio_t}, \text{bio_d}) \in \mathcal{S} \subset \mathcal{N}$ $ (\text{string}, \text{bool}, \mathcal{T}, \mathcal{F}, \emptyset, \emptyset, \text{bio_t}, \text{bio_d}) \in \mathcal{K} \subset \mathcal{N}$ <i>where the string is the node name, the boolean define the abstract state and $\mathcal{T}, \mathcal{F}, \mathcal{S}$ and \mathcal{K} are respectively sets of attributes, flags, sites and key residus.</i> <i>Notice that $\mathcal{S}, \mathcal{K}, \mathcal{A}$ are independent subsets of \mathcal{N}</i>
attribute	::=	$(\text{string}, t : \text{value})$ <i>Where the first string is the attribute name and value is of type "t" $\in \{\text{bool}, \text{string}, \text{int}, \text{float}\}$</i>
flag	::=	$(\text{string}, \text{formula}, \text{formula})$ <i>Where the first string is the flag name</i> <i>The first formula is the triggering formula and the last is the untriggering</i>
action	::=	$(\text{string}, \text{bool}, \text{formula}, \text{formula}, \text{bio_t}, \text{bio_d})$ <i>The first string is the action name, the boolean is the abstract type</i> <i>The first formula is the triggering formula and the last is the untriggering</i>
edge	::=	(in, out) <i>where $\text{in} \in \mathcal{N}$ and $\text{out} \in \mathcal{R}$</i> $ (\text{in}, \text{out})$ <i>where $\text{in} \in \mathcal{R}$ and $\text{out} \in \mathcal{N}$</i>
bio_t	::=	$e \in \{\text{defined biological types}\}$
bio_d	::=	\mathcal{L} <i>a set of biological informations in string format</i>
formula	::=	$\text{true} \mid \text{false} \mid \text{string} \mid \text{att}_1 \text{ comp att}_2 \mid \epsilon$ $\mid \text{formula} \wedge \text{formula} \mid \text{formula} \vee \text{formula} \mid \neg \text{formula}$ <i>Where att_1 and att_2 are attributes with values of the same type</i>
comp	::=	$o \in \{\leq, \geq, <, >, \neq, =\}$

Data structure

- SQL
- Triples ?



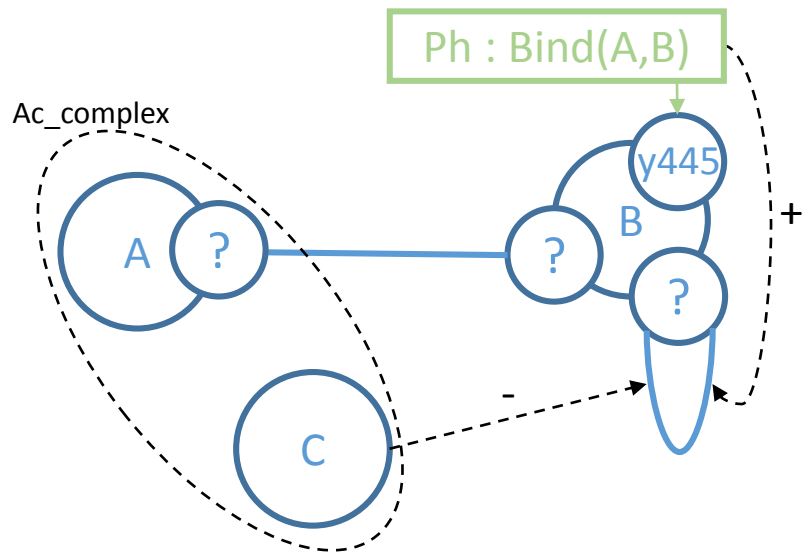


Knowledge inference

Knowledge inference



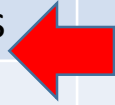
2 kind of knowledge	
mechanistic rules	Assertions
A bind B	C has an influence on B



Knowledge inference

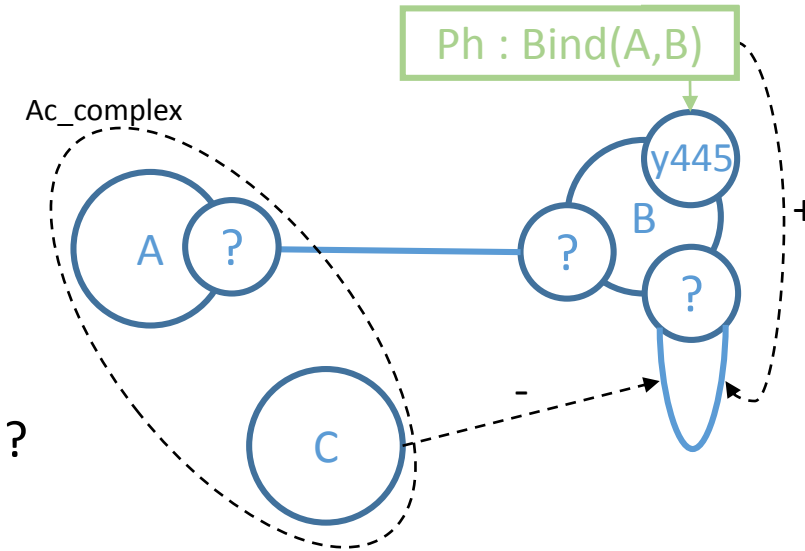


2 kind of knowledge	
mechanistic rules	Assertions
A bind B	C has an influence on B



Problematic :

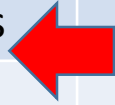
- How to automatically infer formal rules from assertions ?



Knowledge inference



2 kind of knowledge	
mechanistic rules	Assertions
A bind B	C has an influence on B

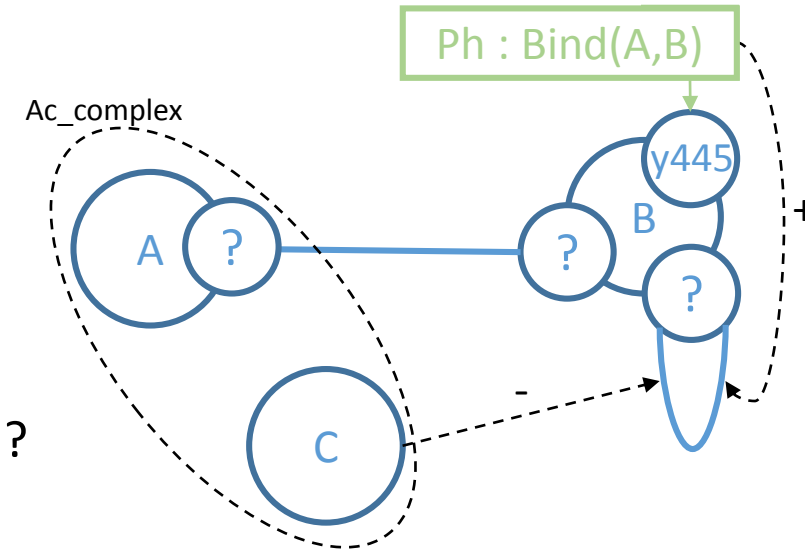


Problematic :

- How to automatically infer formal rules from assertions ?

Approach :

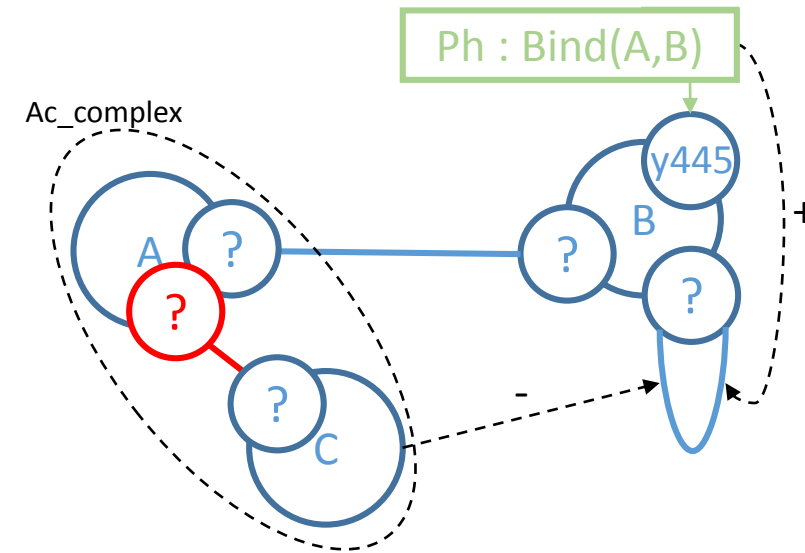
- Looking for pathways into the annotated site-graph



Knowledge inference example



mechanistic rules	assertions
C binds A	AC_complex
B.Bind B ² : B.ph & B ² .ph B.Bind B ² : B.ph B.Bind B ² : B ² .ph	Ph ++> B.bind(B ²)
	C --> B.bind(B ²)



1. Phosphorylation of B has a positive influence on its homodimerization
2. C is a known inhibitor of B homodimerization
3. C is often found in complex with A

The diagram illustrates a protein complex with several subunits. A dashed oval labeled "Ac_complex" encloses subunits A, a subunit with a question mark, and subunit C. Subunit A is connected to a subunit with a question mark, which is in turn connected to another subunit with a question mark. This second subunit is connected to subunit B, which is labeled "y445". Subunit B is connected to a subunit with a question mark, which is connected to a subunit with a question mark. A green box labeled "Ph : Bind(A,B)" points to the "y445" subunit. A dashed arrow labeled "-" points from subunit C to the subunit with a question mark connected to subunit B. A dashed arrow labeled "+" points from the subunit with a question mark connected to subunit B to the subunit with a question mark connected to subunit C.



Merci

Russ Harmer



Big Mechanism



Walter Fontana

